(NASA-CP-2333) TRENDS IN PLANETARY DATA ANALYSIS. EXECUTIVE SUMMARY OF THE PLANETARY DATA WORKSHOP (National Aeronautics and Space Administration) 46 p HC A03/MF A01 CSCL 03B G3/91

N84-34363

Unclas 20508

NASA Conference Publication 2333

TRENDS

IN

PLANETARY DATA ANALYSIS

Summarized by,
Nancy Evans,
Jet Propulsion Laboratory
Pasadena, California

EXECUTIVE SUMMARY OF THE

PLANETARY DATA WORKSHOP

NASA
National Aeronautics
and Space Administration
Scientific and technical
Information Branch
1984

TRENDS IN PLANETARY DATA ANALYSIS

Table of Contents

1. Introduction 1
   1.1 Introductory Statement 1
   1.2 Purposes 3

2. Planetary Data and Planetary Data Analysis 4
   2.1 General State of Planetary Data 4
   2.2 Planetary Data and Its Analysis 6
      2.2.1 Imaging 6
      2.2.2 Non-Imaging Remote Sensing 7
      2.2.3 In-Situ 8
      2.2.4 Radio and Radar 9
      2.2.5 Earth-based Observation 11
      2.2.6 Laboratory Measurements 12
   2.3 The Future Outlook 13

3. Assessment of Scientist's Needs 14
   3.1 Data Needs 14
      3.1.1 Content 14
      3.1.2 Data Access 15
   3.2 System Needs 16
      3.2.1 Central Services 16
      3.2.2 Computational, Output, and Display Requirements 18

4. Technology - Today and Tommorrow 20
   4.1 Computers and Workstations 20
   4.2 Mass Storage Devices 21
   4.3 Communications 22
   4.4 Standards 23

PRECEDING PAGE BLANK NOT FILMED

5. A Planetary Data System 26
5.1 Customer Profile 26
5.2 Functional Description 27
5.3 Physical Description 28
5.4 Pilot Planetary Data System 32

6. Implementation Plan 34
6.1 Phased/Modular Development 35
6.2 Continued Scientist Involvement 35
6.3 Mission Interface 36
6.4 Prioritized Accomplishment 37
6.5 Strawman Schedule 39

7. Conclusion 40

PLANETARY SCIENCE DATA WORKSHOP
NOVEMBER 29 - DECEMBER 1, 1983

WORKSHOP CHAIRMAN

HUGH H. KIEFFER
U. S. Geological Survey
Flagstaff, Arizona

SPLINTER GROUP CHAIRMAN

DATA DEFINITION GROUPS

IMAGING
R. STEPHEN SAUNDERS
Jet Propulsion Laboratory
Pasadena, California

NON-IMAGING REMOTE SENSING
JOHN PEARL
Goddard Space Flight Center
Greenbelt, Maryland

IN-SITU
LARRY H. BRACE
Goddard Space Flight Center
Greenbelt, Maryland

LABORATORY DATA
ROGER H. CLARKE
U. S. Geological Survey
Denver, Colorado

RADIO/RADAR SCIENCE
RICHARD A. SIMPSON
Stanford University
Stanford, California

EARTH-BASED OBSERVATIONS
WILLIAM A. BAUM
Lowell Observatory
Flagstaff, Arizona

HARDWARE GROUPS

IMAGING WORKSTATIONS
LARRY BOLEF
Washington University
St. Louis, Missouri

NON-IMAGING WORKSTATIONS
RAYMOND J. WALKER
University of California
Los Angeles, California

NETWORKING
WILLIAM D. SMYTHE
Jet Propulsion Laboratory
Pasadena, California

SOFTWARE/DATA ANALYSIS GROUPS

IMAGING
LAURENCE A. SODERBLOM
U. S. Geological Survey
Flagstaff, Arizona

NON-IMAGING
RICHARD ELPHIC
University of California
Los Angeles, California

DATA BASE MANAGEMENT
RANDAL DAVIS
University of Colorado
Boulder, Colorado

SYSTEM DESCRIPTION

FACILITY CONFIGURATION
RAYMOND E. ARVIDSON
Washington University
St. Louis, Missouri

FUNCTIONAL REQUIREMENTS
A. IAN STEWART
University of Colorado
Boulder, Colorado

1. Introduction

The Planetary Data Workshop was hosted by NASA, Solar System Exploration Division (Code EL), at Goddard Space Flight Center from November 29 through December 1, 1983. More than 160 planetary scientists joined together, rolled up their sleeves, and went to work. They tackled the problems they felt to be hindering planetary data analysis and developed recommendations to alleviate those problems.

The initial workshop structure consisted of a chairman and 14 chaired splinter groups. Splinter topics covered six data definition areas, six technological areas, and two addressed the system configuration and operations from the user's point-of-view. The workshop structure, including the key participants, appears on the fly-leaf.

Prior to the November session, splinter chairmen and group members worked to prepare strawman concept documents for each splinter topic. These formed the basis for discussion and participation by the workshop body. Synthesis and synergism resulted. Groups were combined; other groups met in joint sessions. Ideas were exchanged, altered, broadened, and new concepts formed.

In the period since the workshop, the resulting group documents have been finalized. The proceedings of that Workshop constitute a Functional Requirements Document for implementation of a Planetary Data System. This Executive Summary provides an overview of the more important aspects of that document.

1.1 Introductory Statement

Planetary scientists believe that two planetary data problems must be addressed: 1) Important data sets are being irretrievably lost, and 2) Use of planetary data is constrained by inaccessibility and a lack

of commonality. Most feel that the present system is inadequate and immediate changes are necessary to insure retention of and access to these and future data sets.

Past successes of the NASA Planetary Program have provided a wealth of data. Each data set is unique as to time, content, format, and means of interpretation. Many of these data are being lost due to deterioration and the lack of an organized system for curation. Once lost, information cannot be regenerated as computers have been replaced and scientists and support personnel have retired or left the field.

Documentation, calibrations, and rationale for making observations were not always preserved. In light of current technology and new analysis philosophies, many of the archived products are not readily useful.

Advances in computer and communications technologies have opened the door to affordable and sophisticated data manipulation techniques. Few, in the planetary community, have had resources to avail themselves of these technologies. There is wide realization, however, of the advantages to be gained: more rapid analysis; analysis of a larger fraction of returned data; and wider distribution of raw and processed data. These new tools can make cross disciplinary studies and the observation of detailed phenomena, over long periods (beyond the span of a mission), a reality.

The evolutionary, technological changes which have caused many of the frustrations in dealing with the current planetary data set also hold the key to alleviating those same frustrations. The current state of much of the planetary data prevents immediate application of newer technologies for data analysis. In order to modernize the system, one must also upgrade and reorganize the data sets. It is possible to develop a Planetary Data System (PDS) which takes advantage of

emerging technology, avoids perpetuating mistakes in the curation of new mission data, and restores essential data sets for productive and continued analysis.

1.2 Purposes

The purposes of this document are to:

1) Outline the current state of planetary data and its analysis.
2) Assess the needs of scientists to perform data analysis.
3) Evaluate the state of technology today and in the near future.
4) Present user requirements for the implementation of a Planetary Data System.

2. Planetary Data and Planetary Data Analysis

2.1 General State of Planetary Data

Workshop data definition activities were performed within six instrument/discipline groups: Imaging, Non-imaging Remote Sensing, In-Situ, Radio and Radar, Earth-based Observations, and Laboratory Measurements. These groups were tasked with defining the state of the data set as it exists today, for each discipline, and discussing means by which the data could be better arranged and accessed to enhance data analysis and accommodate future missions data returns.

Results of the deliberations from these definition tasks were alarmingly similar. Data from recent missions (for which there is currently data analysis money) are locatable, fairly well archived accessible to those who are aware of their existence, but not readily useful to researchers other than the investigation team. Data returned from earlier missions become less locatable, less accessible, and less usable in direct proportion to their age.

Most groups agreed that the currently archived product, with its content and format determined by cost or space constraints and often submitted only to meet mission obligations, is not the most useful in light of new trends in data analysis. The more useful data reside in scattered institutions and government warehouses, curated inadequately or not at all. Large blocks of these data have not been located. Supporting engineering data, instrument calibrations, and documentation, for even recently obtained data sets, are incomplete or missing. More disturbing is the recent realization that reading older (6 yrs or more) data from magnetic tapes can be a destructive process.

The lack of standards for encoding, processing, and archiving data has resulted in chaos for the data user. Not only are data difficult to locate, but they are hard to obtain and laborious to use and

understand. Additionally, the innate instrument differences, time and spatial resolution variations, and nonstandard encoding formats make application of the newer data manipulation technologies difficult.

The opinion that the current NSSDC (National Space Science Data Center) is inadequate for curation, access, and distribution of planetary data is prevalent among data users and data providers. Data archived there are of variable quality, some data sets have questionable accuracy, others are incomplete or have a much lower temporal and spatial resolution than the original instrument data. Little processing history is provided with the data, and engineering information (when available) is supplied as hardcopy. Reprocessing or reanalysis of most data obtained through NSSDC is not possible, especially since much of the archived data is in analog format: pictures, graphs, plots. The NSSDC data catalog represents the total of the data resident at NSSDC. It, however, does not catalog data to a level sufficiently deep to identify specific observations of interest.

The reasons to explain the woeful state of planetary data are many. These major reasons were discussed at the workshop:

> Both science data analysis and preparation of data for submission to archive are included as one line-item in the mission science budget. At times of severe budget constraint, when there is insufficient funding to support either task adequately, the motivation to perform data analysis is stronger than that to organize data.
>
> Additional funding is not provided to NSSDC to organize data which are inadequately submitted or to process or manipulate these data sets to improve usefulness. The highly useful data sets in the NSSDC are those which were acted upon by scientists and the NSSDC after original submission.

2.2 Planetary Data and Its Analysis

Planetary data analysis is the application of theories and algorithms to data (raw, reduced, or preprocessed by someone else into a more usable format) to generate information about the Solar System. Investigators in the field are divided into many specialty groups or disciplines. Research facilities vary in size from small (a single scientist in a college or university) to very large (a university department or a branch of an agency). The ways in which these investigators use their data vary as radically as the types of data analyzed and the computational equipment required.

Disbursement of analysis tools is directly related to one's success in proposal writing. Availability of digital data to scientists other than team members is a limiting factor. No institution is chartered to archive or distribute digital data. Data of questionable quality and pedigree are available from the NSSDC. These data are (by charter of the center) reduced data products; many are no longer in digital format. Regional Planetary Imaging Facilities (RPIFs) provide data access for that community through browse and catalog search functions. They do not support distribution of hard-copy nor digital data tapes. Therefore, much of the data analysis is performed upon "boot-legged" versions of a public domain commodity.

As each data type supports analysis by various science disciplines, data analysis will be briefly discussed within the framework of data class.

2.2.1 Imaging

Imaging data include those measurements made by Vidicon, Reticon, and (soon) CCD camera systems aboard planetary spacecraft. Spectral information, in the visible and near UV wavelengths, is obtained through the use of filters. The raw photometric, digital data are

processed and enhanced into photographic image format. Both digital and hard-copy image data are used by the imaging scientist in analysis. Imaging data support analysis by two major groups: atmospheric scientists and geo- or surface scientists.

The more closely the discipline is dependent upon describing what something looks like: its boundaries, its shape, its irregularities, its subtle differences from other similar objects, the more reliance is placed upon hard-copy images. These are also used by other instrument groups to locate an observation or substantiate a boundary or other unusual finding.

Geo-scientists require both hard-copy and digital products to support analysis activities. Photogeologic analysis is done by making visual comparison of terrain under a variety of lighting and viewing conditions. Many observed photographic enlargements and mosaics are used in this comparative technique. Digital data are then used to make geometric and photometric measurements to support these visual analyses.

Atmospheric scientists and geophysicists perform more digital processing than the group of scientists discussed above, and have greater requirements to combine the imaging information with measurements from other instruments. Both groups tend to analyze data in the temporal as well as the spatial domain and place heavy requirements upon the accuracy of the accompanying navigation and engineering data. Atmosphericists, in particular, make heavy use of browse capabilities (ideally from a CRT display) to enable identification and comparison of cloud features.

2.2.2 Non-Imaging Remote Sensing

These data include spectrometric, radiometric, and polarimetric remote sensing observations in wavelengths between gamma-rays and microwaves. High resolution measurements (better that 1% of the

wavelength) are useful for identifying atomic, molecular and ionic species; broader band measurements are adequate for identifying minerals and determining total energy fluxes; polarization data are used in the study of finely divided material such as clouds and surfaces. Users of remote-sensing data, like users of the imaging data, are divided into two camps: those who study atmospheres and those who study surfaces.

Analysis of remote sensing data, perhaps because of the many instruments represented or the wide range of wavelengths measured, places the heaviest requirement (of all data types included in the study) upon being able to overlay and compare measurements. Of specific concern are those measurements made at the same time and the same place by several instruments or made of the same object at different times by the same or very similar instruments. The remote sensing scientist compares the observed against the known, the modeled, and the average to yield a difference. Differences are then represented graphically.

This group is affected most seriously by differences in temporal and spacial resolution, by incompatible data formats, and inadequate or incorrect pointing parameters. They are most seriously frustrated by the inability to easily compare data obtained from different instruments and from different missions.

2.2.3 In-Situ

In-situ planetary data refers to those measurements made by a wide variety of instruments whose sensors are exposed directly to the planetary environment. Instruments may be mounted on a fly-by or orbiting spacecraft to monitor conditions along its trajectory or on a lander or probe to monitor the descending or surface environment.

Typical measurements include magnetic fields, energetic particle densities and energy distributions, plasma wave characteristics, ionospheric and neutral gas densities, temperatures, compositions, and motions.

These measurements support data analysis in the study of atmospheres, magnetospheres, and the solar wind phenomena. Interdisciplinary investigations involving the solar wind interaction with planetary magnetospheres (Jupiter, Saturn, and Earth) and with planetary atmospheres and ionospheres (Venus, Mars, Comets, etc.) require in-situ data from a wide range of instruments, different spacecraft, and different planets. Because of the interdisciplinary nature of the data set, data users place heavy demands upon each other and upon the coorelative data sets, for specific temporal periods, that have been generated at NSSDC.

Analysis of in-situ data is performed most frequently on data which have been processed to physical parameters of the environment. It is felt by this group that most data users do not have the expertise nor the computing capability to convert the raw data to useful form. However, with this in mind, this data group most strongly of all advocates preservation of raw data indefinitely. More recent discoveries about the planetary environment or the operation of an instrument, either from a later mission or later in an ongoing mission, have already required reprocessing of the raw data to extract additional information or to improve the accuracy of reduced data. Currently, these raw data are archived by the instrument groups themselves.

2.2.4 Radio and Radar

Radio science experiments use electromagnetic waves to probe or study the solar system. Three major disciplines are supported by this data set: radio astronomy, radar astronomy, and celestial mechanics.

Radio astronomy is the detection and measurement of naturally produced radio frequency emissions from planetary surfaces, atmospheres, rings, and plasmas. Results can be expressed in "bulk" properties, or as maps if the source can be spatially resolved. Maps are a common output of interferometric (multiple antennae) observations. Although most radio astronomy is conducted from Earth-based facilities, two spacecraft--Pioneer Venus and Voyager--carried radio astronomy instruments.

Radar astronomy involves the observation of man-made signals after their interaction with a target. Targets include surfaces, atmospheres, rings, and plasmas. Both imaging and non-imaging data may result and can be presented in a time domain, a frequency domain, or both. Monostatic radar observations are made with a single antenna acting as both sender and receiver (Pioneer Venus and the Earth-based Arecibo and Goldstone studies). Bistatic radar utilizes separated antennae for sending and receiving. Usually, the receiver is on the ground, while the transmitter is carried aboard a spacecraft.

Celestial Mechanics includes all studies related to the motions and gravity fields of bodies within the solar system. Spacecraft tracking radar ranging, observations of relativistic effects, gravity wave studies, and radio occultations are included in this category.

Unlike most of the other instrument data, the usefulness of radio science data is increased in direct relationship to the level of processing. However, the derived data is virtually unobtainable through NSSDC while large quantities of raw and calibrated data are archived there. Locally archived sets of derived data and data sets under investigation are to be found in the investigators institutions.

Analysis of data from radar and radio astronomy experiments yields information regarding the density (profile) in atmospheres, particle

sizes in planetary rings, or particulate and roughness characteristics of target surfaces. In situations where the entire body of the object is the target, density and rotation rate are derived by analysis. Often, data obtained in ground observations are combined with those obtained from spacecraft to increase coverage or resolution in the analysis procedure.

Analysis of Celestial Mechanics experiment data supports the characterization of gravity fields and, thence, the determination of planetary masses and mass distributions. The data are also used in the production and maintenance of ephemerides which, in turn, support mission navigation and sequencing activities.

2.2.5 Earth-based Observations

Earth-based astronomy data complement spacecraft data by providing observational coverage of important planetary phenomena at epochs before and after spacecraft encounters. These data have longer time scales, and commonly have broader wavelength coverage and higher spectral resolutions than spacecraft data. Included are observations made by ground-based telescopes; aircraft, balloons, and rocket-borne instruments; and Earth-orbiting spacecraft.

Observations grouped under this category are similar to those included in both the imaging and the non-imaging remote sensing areas (the exception being the large volume of analog data in the form of photographic plates). Measurements are manipulated and analyzed in the same way as spacecraft data; often by the same group of scientists. The primary importance of this data set to space exploration, planetary mission planning, and the analysis of spacecraft data (and the reason for its inclusion), is that it provides a history and a continuum from which to draw broader-based conclusions. It also provides a basis for selection of instruments to be included on exploratory spacecraft missions.

Major problems exist in the acquisition, use and/or analysis of these data. A large percentage of the ground-based observations is made by individual investigators acting largely on their own initiative. Often such work is supported by federal grants and the investigator is obligated to publish his results in a scientific journal but is not required to archive the data in the public domain. Earth-orbiting observations are organized with some similarity to planetary missions and requirements are levied to submit data for archive. However, format variations exceed and cataloging and documentation deficiencies are comparable to those of the planetary data set. Many of the older, and therefore historically useful, data exist only in hard-copy format and must be digitized when used for more than visual comparison.

2.2.6 Laboratory Measurements

Laboratory data are traditionally the measurement of one parameter as a function of one or several other parameters: intensity, transmission or reflectance versus wavelength; intensity versus wavenumber; band strength versus temperature; sputtering yield versus impact energy; etc. These values or standards are acquired in numerous research laboratories around the country and world, and are vital to NASA Planetary Programs as they provide the basis for the interpretation of planetary data.

Data provided to the planetary community to support analysis are primarily paired (x,y) data. Some data are tabular in nature but can be used as a multiple, paired data set. An extension of the paired format concept is three-dimensional data, which can be treated during analysis in a manner similar to the paired data.

As with the Earth-based supporting data, laboratory data are obtained by individual investigators who are funded by organizations other than NASA to perform the research activities. The investigator is required to publish the results of his research; thus the methods and

measurements are well documented. However, no requirement is placed upon archiving data in the public domain. Many of the more active researchers maintain personal or institutional archives of spectra. For some fields of research, particularly spectroscopy, there is a community attempt to standardize data formats.

2.3 The Future Outlook

The emergence of the Interdisciplinary Scientists (mission science members whose research responsibilities include the output from many instruments on the spacecraft) and the discovery of elements in the solar system too complex to be understood by isolated analysis attempts--Saturn's Rings, the Venus environment--have created new requirements for effective data analysis by the planetary community.

The desire to compare one data set with another, at the highest possible resolution; to view in synopsis, temporal or spatial, all observations of an object or phenomena; to transfer raw data bits or analyzed results rapidly between investigators; and to coordinate results and reexamine details in real time; place requirements upon the data set and the current data analysis system that cannot be met. The archaic system cannot support modern data analysis techniques. The situation is not only frustrating to the scientist, but self-defeating of NASA's goal "to be the world leader in space exploration."

The ensuing sections of this report summarize the requirements of this group to enable performance of the planetary data analysis task. Discussed are a Planetary Data System: the services it should provide to the user, how it should function as a system, and the technologies which can be employed to accomplish those ends. The viewpoint is now. The future extends as far as analysis of current technology trends permits.

3. Assessment of Scientist's Needs

The building of a Planetary Data System (PDS) is predicated upon meeting the needs of the planetary scientist to perform data analysis. The single activity of greatest importance at the workshop--and in which all individuals participated--was the definition of those needs. Individual needs were assessed as to their importance within the general structure. Then, needs from all interest groups were coordinated and integrated to form a set of blanket requirements which should be met by the proposed system.

These have been separated into data needs and system needs for discussion.

3.1 Data Needs

3.1.1 Content

There is general agreement that the PDS should function as both a tool for active research and a permanent archive. It should contain not only instrument data and data products, but also the engineering and pointing data which make the measurements meaningful. Thorough documentation of the data, the instrument, and the mission sequence objectives must be included for each data set, and if analysis and/or reduction software is supplied, it must be documented as to its use and reliability.

The need to retain many levels of data to support the requirements of several kinds of users was recognized by all data groups. If the PDS is to meet its design objectives, the data complement will include raw (decommutated) data, processed and partially processed data resulting from the application of reduction and analysis procedures, and finally, the information generated through scientific analysis. As the level of processing increases, data should be more tightly screened for reliability and accuracy. In addition, a method

for updating data in the archive must be incorporated, based upon improved information, and including additional parameters or associated variables. Current developments in data compression and high density storage make archiving such a vast volume of data possible.

3.1.2 Data Access

Interactive access to the data sets themselves, though of great importance to data analysis, was deemed less important than access to a sophisticated catalog from which to identify observations and measurements of interest, browse through candidate data sets, and place orders for selected data and information. An on-line catalog, allowing entry at several levels, was the foremost requirement to be placed upon the PDS. The catalog should, therefore, be considered the heart of the system, and its development must be the initial activity in the modular PDS implementation plan.

The catalog will be the initial entry point for all levels of users into the data system. Characteristics, upon which all agree, are that: the catalog be on-line and interactive; there be entry levels into the system which vary from system novice to system expert; level of access, security, and use of services should be controlled by "smart" sign-on procedures; all data in the archive and all system services should be identifiable through this mechanism.

Accessibility through traditional means, i.e. visits to discipline centers, the Regional Planetary Image Facilities, and PI institutions--with one to one interactions with scientists, were considered important. Retention of personal relationships in data analysis was strongly advocated. The concept of Discipline Centers, discussed in section 5, fosters the interaction and personal relationships between scientists.

3.2. System Needs

Workshop participants were less uniformly aligned behind the system requirements than behind the data requirements discussed previously. A several-layered structure was indicated, which included centralized and distributed functions. Generally, all agree that the initial access catalog should be a centralized function and that the second layer, Discipline Centers, should be distributed from the central function. However, each Discipline Center would retain the option to develop either a centralized or a distributed system to support discipline-specific data analysis tasks. A combination of Investigation Nodes and Analysis Nodes make up a third layer within the distributed Discipline Center concept. A diagram of the proposed functional architecture is included in Section 5., Figure 5.2.

Functions suggested for incorporation into the PDS are strongly dependent upon a communication and data transfer network. It is to be assumed for purposes of planning that the Code T, PSCN (Program Support Communications Network), as presented to the workshop, is indeed a reality. The creation of a separate network to support Planetary Data Analysis is not an affordable option. The use of commercially available network services seriously constrains the scope of the PDS.

3.2.1. Central Services

A user must be able to determine easily whether data useful for his purposes exist; examine them in increasing detail; and place an order for transmission of those observations or data sets which are appropriate. Identification, selection, and distribution of supporting information (documentation, calibration files, engineering information, and software procedures) will be included in the central service function. While this is a continuous process, there are seven distinct stages or support services included; the first three of which constitute the catalog.

A. Encyclopedia - is a general guide to planetary data. Organization in four categories, hierarchically related, is indicated:

1) Solar system components
2) Missions and spacecraft descriptions
3) Instrument descriptions and objectives
4) PDS services and facilities

B. Data Index - lists and describes the data and information contained in the archive. The description for data from each investigation should include an instrument overview, temporal and spatial coverage, levels of data products available with data quantities and qualities indicated, and resource contacts.

C. Data Inventory - expands the descriptions in the Index to allow an investigator to pinpoint the observations appropriate for analysis. Suggested for inclusion were cartographic coordinates, resolutions, instrument states, etc.

D. Browse - permits examination of data at a level of detail equivalent to leafing-through hard-copy. The user should be able to designate a step-function, halt the progress to observe detail, and resume the function. The capability to specify data of interest to be extracted with the sample or acquire capabilities should be supported by the browse function.

E. Search - allows search and retrieval of data or information which satisfy specified conditions. A function which permits the identification of overlapping observations (temporal, spatial, or target specific) is required to support interdisciplinary data analysis.

F. Sample - provides a representative small sample of a data set which may either satisfy a user's needs or may be used to test material for development of analysis procedures.

G. Acquire - enables user requests for data to be transmitted later (either electronically or by postal or package services as determined by data volume and urgency of the request). The first response to any request should be an estimate of cost (if any) and time delay associated with filling the request according to different modes of delivery. The acquire function will also support the distribution of mission data sets to investigators.

3.2.2 Computational, Output, and Display Requirements

The ability to access the data directly and to manipulate and/or analyze the data is a requirement of a PDS. Several kinds of basic data processing are required. These are:

Manipulation - simple algebraic or trigonometric operations to generate a new parameter from one or several data parameters.

Statistics - reporting the statistical properties of one or more parameters over a spatial or temporal range.

Registration or resampling - interpolation or changing of temporal or spatial resolution of data.

Display - format change and graphic presentation of data values.

Analysis - operations more complex than manipulation which often require application of scientific models.

Recalibration - processing of raw data with altered calibration files or procedures.

Access to several classes of computers, high-speed printers, and expensive display and reproduction devices will be provided through the network. The function must appear central, though the location of devices may be distributed. Scientists may augment their home institution capabilities by interactively accessing one or several of these devices, or tap into the services from a Discipline Center.

4. Technology - Today and Tomorrow

The technology revolution is providing the building blocks for a Planetary Data System. Market researchers and financial analysts indicate that the technology explosion has just begun. However, the same technologies which enable a PDS, will inhibit its function if heavy dependence is placed, today, on hardware- and software-specific implementations.

Assimilation of rapidly advancing technologies into plans and decisions is difficult. Dreams have become realities overnight. On the other hand, the incompatibilities fed by fierce competition are creating large problems for systems, such as PDS, which must accommodate the past into the present while, at the same time, planning for the future. The task is not unique; the entire world faces the same problem, i.e., coming to an agreement on the way to manage information while the information industry improves technology faster than components can be incorporated into plans. This is an enormous order for a world which currently supports incompatible telephone systems, several railroad gauges, and can't decide on which side of the road to drive a car. The PDS should isolate into easily changeable modules those functions that depend upon specific hardware or software implementation.

4.1. Computers and Workstations

The use of mainframe computers and shared access mini-computers for research and analysis computation is well understood by the planetary community. Miniaturization is reducing the size and the cost of the mini-computer at approximately the same rate that mainframes increase speed and performance capabilities. Digital Equipment Corporation's VAX, the basis of many planetary data analysis systems, now is being marketed as a desk-top unit--not as powerful as some models perhaps--but that day is coming.

At the same time, the personal computer revolution has encouraged manufacturers to design ever more powerful microprocessors. The more sophisticated systems support high density disk storage, mini-level languages and procedures, and bit-mapped graphics displays in monochrome and color. The "souped-up" micro, now, can form the heart of an inexpensive yet powerful scientific workstation, and this trend of expanding capabilities can be expected to continue far into the future.

The analysis systems required for planetary science users must support the following functions for all operations, without regard to the complexity of the task.

- o Provide access to data
- o Process data
- o Store data
- o Display data

The magnitude of the analysis task and the complexity of the operations required determine the class of equipment which should be used to accomplish those functions. If computational needs for analysis exceed the capacity of the user's system, larger scale computational facilities can be made available on a time-share basis through network connection.

4.2. Mass Storage Devices

The technology area of most rapid growth is that of mass storage. The density of encoded data bits upon magnetic substrate, either tape or disk, has quadrupled in the past five years. Standard 1/2 " 9-track magnetic tape is now encoded at a density of 6250 bits per inch (bpi); 1600 bpi is still standard on most systems used by the planetary community.

Not only has the capacity increased, but the physical size of the unit is shrinking. In the world of micros, a ten megabyte hard disk drive now occupies less space than the traditional 5 1/4" floppy disk drive. More powerful systems drive larger capacity disks. Floppy disks now are encoded in quad-density and some, for the newest systems, equal the capacity but measure 3 1/8" across. One of the newer microprocessors incorporates 384K bytes of bubble memory into a 16 bit system and is one-half the size of an ordinary briefcase.

Requirements for archival storage and distribution of large quantities of data will be met by the analog videodisk, digital audio disk, and write/read optical disk storage systems. The most attractive of these new storage media for planetary use is the optical disk which offers the potential of storing gigabytes of data on a single disk--the last 20 years of planetary exploration on 100 platters. Test units are being distributed now and commercial units are expected on the market in the very near future. The life-expectancy of the optical disk far exceeds that of any digital storage media available today.

4.3. Communications

Communications networks sprawl across the globe, some bouncing from surface to satellite and back to the surface. The Era of Information is upon us. Efficiency means rapid transfer of ideas, data, decisions. Gone are the days of the pony express and the telegraph. Even the three-hour delay between the start of the working day on the East Coast and the start of that same day on the West Coast--and the converse--seems less impactive with the advent of electronic mail; one can leave a message to be acted upon.

Satellite communications and packetized telemetry technologies provide a high-speed pipeline for transmitting large quantities of data, rapidly and accurately, from computer to computer. Proprietary networks which support mail services between users, are developing

voice communications. One company is offering this service to a limited number of users for test. Another offers an interactive teleconferencing package which supports 1500 conference members in 35 activity groups, and can handle interactive graphics if all members are using the same microcomputers.

Communications are necessary for carrying on the business of science. Communications technology can be employed in locating and obtaining data, computing, cross-fertilization of ideas, and resolving a wide variety of operational issues. Communications and network technology are vital to support the proposed configuration of the Planetary Data System. The core of the system, that which supports inquiries and catalog services, requires basic communications technology in order to function.

Incompatibilities between classes of equipment, data formats, and system software, are manifest most severely in the area of communications. The adoption of standards for encoding data and using data by the planetary community, will alleviate local incompatibilities. However, the major advances must come from the computer industry and the international standards committees. Commercial information networks serve, now, in the capacity of universal protocol convertors. Small-scale user groups are forced to employ commercial techniques and wait for the world to come to agreement as to how information will be exchanged.

4.4. Standards

Standards are ubiquitous in data processing: the width of computer paper; languages, such as Fortran or Basic, which with a few changes produce the same program on an IBM system as on a DEC system; or the protocol which allows the printer to understand the processor, or the processor the disk drive. Standards are difficult to develop, they require agreement of all parties involved, and this is difficult when considering major issues. Development of a standard operating system

or a standard protocol means "standard" for those making the decision: a computer manufacturer, a company, an organization. "Therein lies the rub." It was never implied that standards should be universal, and up to now it wasn't that important.

The "standard" of greatest concern to the PDS is that of presenting a standard appearance to the user. Any user (with access to appropriate equipment) should be able to address the system, obtain information, and order data and data support products which will behave on his system in a predictable manner to yield predictable results. This implies a level of system compatibility and data transportability that is not currently available. Two very large scale standards efforts are currently underway which will facilitate a standard operating appearance.

These are discussed below. Until those decisions are made, it will be necessary to employ work-around procedures, protocol convertors, and software applications.

Two standards, in particular, will be incorporated into the PDS upon adoption: standard communications protocol and the standard format data unit (SFDU).

1. A standard communications protocol, or conversion of all protocols to a standard, permits easy access by all to the central catalog, the network services, and the archive data bases. Proprietary networks have, for the most part, overcome the machine-dependent incompatibility issue, but the design of the catalog, the data base and the network services must support laissez-faire access by the community.

2. SFDU is defined as a unit of data that has been encapsulated within a globally interpretable primary label. Much work has been done by an International Standards Committee to develop the SFDU, whose purpose would be to provide a means for global

identification of the structure of a data unit. It is currently understood that the primary label contains the control and format codes which permit identification of the data source and encoding format of the data. The creator of data will then design the data unit contents to meet his needs from a standard set of formatting structures.

5. A Planetary Data System

By definition, the Planetary Data System is an interactively accessible data analysis tool for the planetary scientist. It must embrace those functions which planetary scientists feel will aid them in performing better data analysis. Such a system should provide a fertile environment for science while "invisibly" accommodating the ups and downs as science priorities change, technology promotes daily obsolescence, and funding levels for support vary annually.

Following implementation, the business of planetary data analysis will be performed in a different way. The impact of this will transcend the boundaries between NASA Centers, projects, missions, and the individual science disciplines.

5.1. Customer Profile

The original, altruistic concept, that the Planetary Data System provide service to all who need access to planetary data, was revised to reflect a more conservative position. The less knowledgeable user has a significant impact on system design and cost in terms of the introductory help required and the additional entry levels into the system which must be provided. If Public Information and Education (below University level) needs are to be met, then funding must be provided from those quarters to enable the PDS to incorporate support of these functions.

Therefore, the primary user of the PDS is viewed as being an active research scientist performing basic research activities upon planetary data or relating planetary data to an associated research discipline. Extremes in this category include inter-disciplinary scientists, graduate students, and NASA mission planners.

Users will be computer literate, aware of the mission parameters, and familiar with data from at least one type of instrument.

Cross-disciplinary studies will require that data from all instruments be documented to a level sufficient for use by this typical user.

Scientists requiring heavy use of computing capabilities will be located in small private industries or smaller colleges and universities, while the larger institutions are anticipated to be the suppliers of this computing capability. Implementation of the Center of Excellence/Discipline Center concept will require that scientists from those organizations become mentors as well as users.

5.2. Functional Description

The minimum configuration for a Planetary Data System is expected to provide the following attributes.

The PDS should include a working (as well as a historic) archive of many levels of data products and the information necessary to make these useful.

Data and services should be presented through a searchable catalog with means provided to ascertain the appropriateness and the quality of data.

Selected data should be made available through a distribution function.

Computing capabilities to support scientists who have limited facilities or infrequently require greater computing power than is available to them directly should be provided through some on-line or switching function.

Heavy reliance is placed upon communications technology to carry out these functions. Access to the centralized data catalog will be provided through interactive communications from a work station in

the home institution or Discipline Center. Transport of data to the investigator, as the result of a request, will be made by the most appropriate means, taking into account the urgency of the request and the quantity of the data. Small amounts of data, needed in a hurry, will be transmitted directly. Large amounts of urgently needed data might be sent through the packet telemetry channel. Routine distributions or large volumes of data will be encoded upon one of the future high-density disk media and sent by package delivery, certainly the most economical of distribution methods.

Functionally, the system must be easy to use. These important aspects of system operation were discussed during the workshop. Instruction for the novice user must be conducted on line, through the sign-on procedure, while operations for the expert user must not be constrained by reams of menu functions. Transfer from centralized to distributed functions should be invisible to the user. However, if a facility is busy or unavailable, a queueing function should be provided to allow the user to leave the line. All in all, the reaction of the entire system to user loading (not just input and output ports) should be well understood, so that the system remains responsive to requests. Long waits for answers are not only frustrating to the user but expensive from the standpoint of hook-up charges and line time.

5.3. Physical Description

The system configuration, currently considered to best meet the requirements for a planetary data system, combines the attributes of a distributed system with those of a centralized system. Figures 5.1, Preliminary PDS Management Configuration, and 5.2, Preliminary PDS Functional Configuration, present the system from both important points of view.

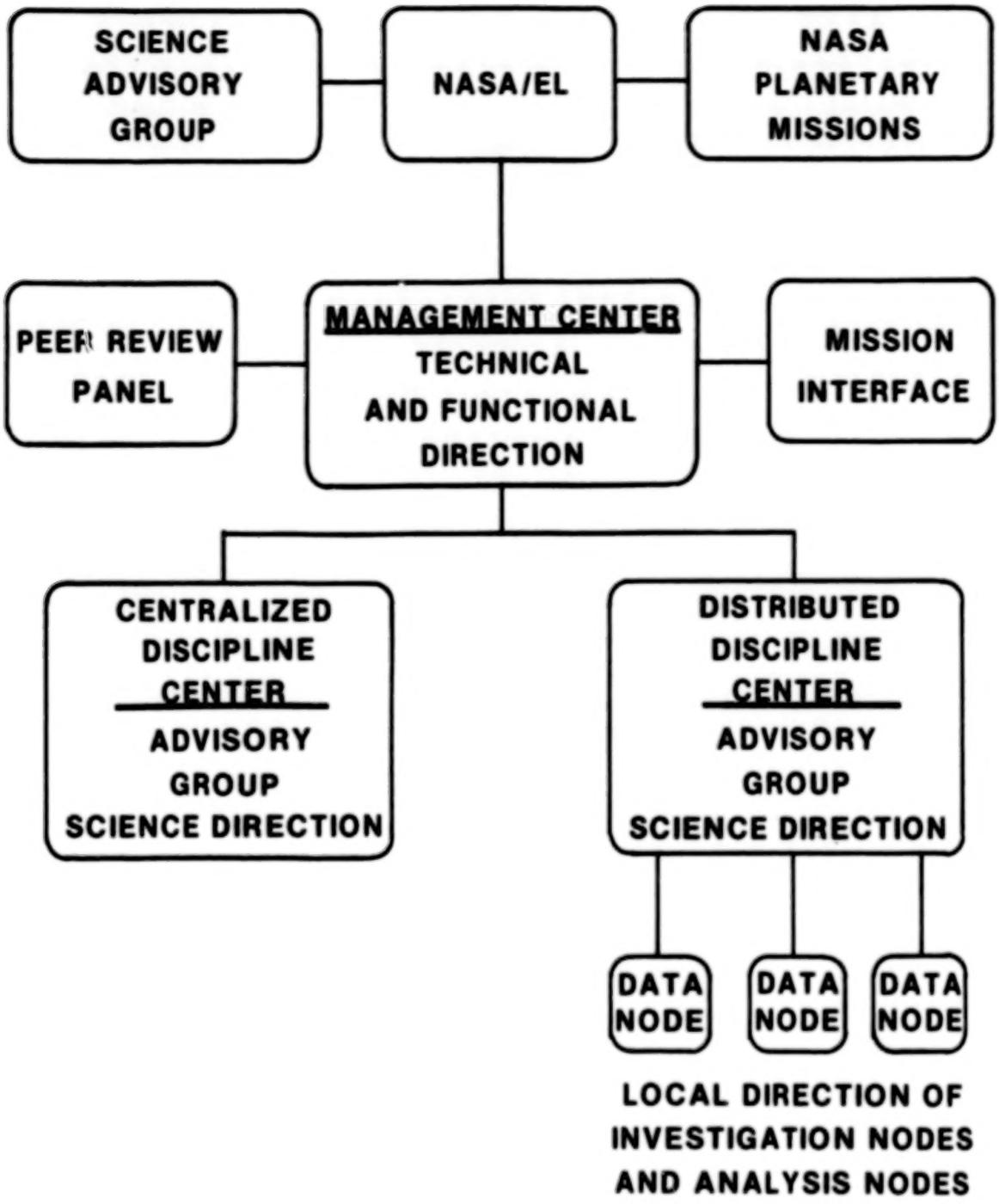

FIGURE 5.1 PRELIMINARY PDS MANAGEMENT CONFIGURATION

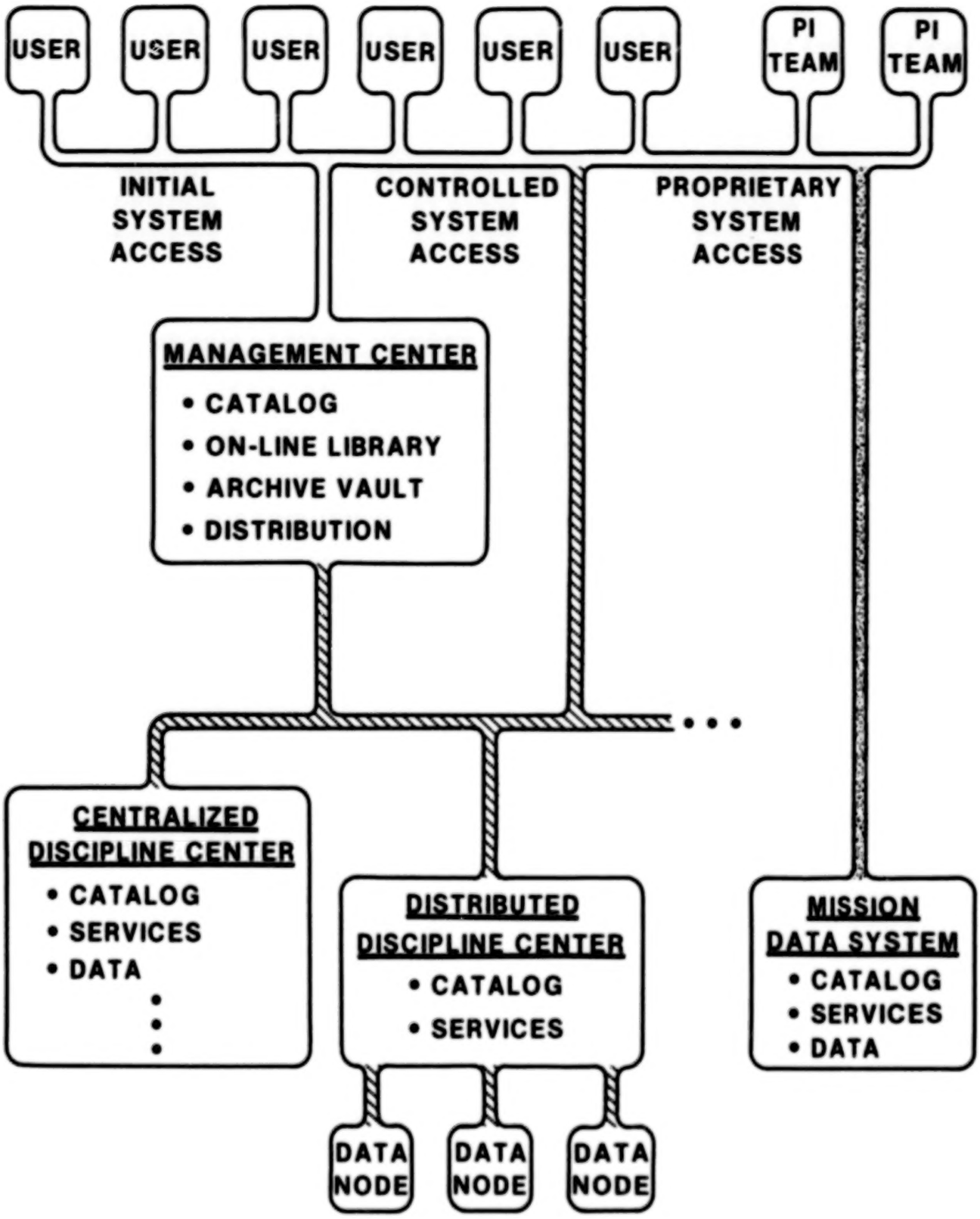

FIGURE 5.2 PRELIMINARY PDS FUNCTIONAL CONFIGURATION

Elements of the Preliminary PDS Functional Configuration are explained in detail in the Planetary Workshop Report. A brief summary, however, adds meaning to the figure.

Advisory Group - a committee of planetary scientists to advise NASA on issues of policy and goals for PDS, and its interactions with the community and flight missions. May initially be the PSDSG.

Management Center - many functions of the PDS are general to all disciplines and should be centralized. Specific among these is database management control, network control, funding management, maintenance of the archive and the catalog, and distribution of mission data submissions to active data base sites.

Peer Review Panel - greater than casual access to the PDS facilities, most especially the requirement for large quantities of data or use of computing and output capabilities will require review by a panel of one's peers. In these cases the PRP will act as an access authority.

Discipline Center - associated with an active research group, is supported by the "Center of Excellence" concept. As such, the center will provide expertise on the applicability of data to research problems, aid scientists in the interpretation and use of those data, and provide a suite of facilities appropriate to perform research in that discipline area. The Discipline Center may adopt a centralized or a distributed configuration. Centralized implies that all data and responsibility reside in the Discipline Center. Distributed indicates that some or all of the data and responsibility reside outside of the center in Data Nodes.

Data Node - One of the principles behind PDS is that data are best kept in an active archive by scientists who understand them. This network site supports that premise. Data Nodes may vary in size and

function from a single individual to an entire department whichprovides many services of the Discipline Center. Two types of data nodes are defined:

Investigation Node - contains the complete data record of a planetary investigation. The prime responsibility for the data set remains with the Principal Investigator or Team Leader (or his institution) until relinquished to the Discipline Center.

Analysis Node - contains a major analysis data set, e.g. a consortia or highly processed and resampled data set. The generator of the data set has prime responsibility for its use and upkeep.

Users - General users of the system will have either primary or controlled access to the PDS according to requirements placed upon the system function and services.

PI Teams - Only currently active PI Teams and Interdisciplinary Investigators will have proprietary access to the network.

Discipline Advisory Group - each discipline will select, on a rotating basis, a group to advise upon the content and the functions of the Discipline Center.

Network - a telecommunications network, linking data sources with data catalogs, and data services with data users, is the circulatory system of the PDS.

5.4. Pilot Planetary Data System

The Pilot Planetary Data System (PPDS), managed by Code EI, Information Systems Office, and advised by Code EL, Solar System

Exploration Division, is now under development at JPL. This pilot is divided into a set of coordinated tasks designed to test certain elements of a system to support planetary data analysis. Developments in the areas of networking, optical disk technology, navigational support data, and database management are being explored and incorporated into several science support tasks. These are:

Planetary Imaging, Mars Surface Properties and Aeronomy, Jovian Magnetosphere, and Planetary Rings.

As elements of a planetary data analysis support system are proven useful by the PPDS, their incorporation into the functions of the PDS are anticipated. Strong ties between PPDS and PDS insure a rich technological and operational inheritance.

6. Implementation Plan

Implementation of the Planetary Data System (PDS) must take into account the many important facets of supplying the planetary community with a highly useful but affordable tool for data analysis. Foremost in the minds of all workshop participants was that the PDS not become what was variously termed: "A Monster System", "A White Elephant", or "A Resource Drain".

The importance of several elements included in the implementation plan is stressed. It is felt that these will assure that the PDS is useful to the community, and is developed in the most cost conservative manner.

1. Existing data sets must be located and upgraded for inclusion in the system and steps taken to ensure that future data submissions are in an appropriate format.

2. The PDS should grow by the incorporation and upgrade of existing analysis facilities into the core structure. Necessary, but costly, equipment should be placed on the network for shared use by many.

3. Sufficient resources should be allocated to support the implementation of the system, as well as to provide scientists with the necessary funding to use it.

4. Planetary scientists should remain involved in all phases of planning, implementation, and operation. The scientist best understands his needs for data analysis and best relates excessive cost with reduced analysis resources.

6.1. Phased/Modular Development

Economic constraints provide the most serious driver to the system design and the period over which implementation will take place. Implementation is, therefore, viewed as a paced growth process continuing over a number of years. A fixed-level of funding is required throughout the developmental period. If the funding level is high, more tasks can be accomplished each year, and the period of development will be shortened. A lower level of funding will result in an longer implementation period.

Establishing priorities between data reclamation and system development causes tremendous conflicts for funding and attention. Saving the data, in jeopardy of permanent loss, has been given the number one priority. However, if the cataloging system, the archiving system, and the standards tasks remain static, the data will remain unavailable. It is hoped that many of the strictly technical issues will be addressed by the Pilot Planetary Data System (PPDS), and progress by the Pilot must be folded into the modular plan. A high level of coordination between the PPDS and PDS efforts is felt to be of the utmost importance, especially since the PPDS approach is to develop many tasks simultaneously, while PDS will, at best, assimilate one development at a time.

6.2. Continued Scientist Involvement

Implementation of the PDS requires the active participation of planetary scientists. The system structure, as it is currently understood, is a network of existing facilities; selected by the peer review process from submitted proposals; and updated through modern technologies to support a wider circle of planetary scientists. Discipline Centers and the various Data Nodes will be formed by this

process. Inclusion of existing computational and data storage facilities, their active data bases and active researchers, makes good sense from a monetary standpoint and alters, least significantly, the way scientists perform data analysis.

Reclamation of data--processing or reprocessing, formatting, documentation--is another area in which committments to participate are required from the community. Through an AO process, Investigator and Analysis groups (or individuals) will have an opportunity to describe the state of the current data set, suggest appropriate enhancements to make data more usable within the guidelines of the PDS, and estimate the costs involved to perform these tasks. Such proposals will provide answers to three important questions:

1) What is the scope and the cost of the data reclamation effort?

2) Who in the community are interested in participating in the data reclamation task?

3) What data sets can be included in the initial PDS configuration?

Scientists will make up the Advisory Committees and the Review Boards. Scientists will determine policy, establish procedures, and decide trade-off positions between usefulness of certain data sets, implementation of new technology, and the budget. Furthermore, interaction of the science community with the system in a custodial manner will be required as long as the system is in existence.

6.3. Mission Interface

Relationships between scientists and flight projects will be altered by the Planetary Data System. Planetary scientists feel that change in this area should be instituted immediately, certainly in time to

influence MGCO (Mars Geoscience Climatology Orbiter). Announcements of Opportunity for flight investigations should describe in detail the requirements for submission of data and documentation to the PDS. Significant variation from those non-specific requirements for submission of data to the NSSDC, currently in use, were recommended. Workshop participants were adamant that a means to ensure PI submissions be instituted. However, most were willing to admit that submission of inadequate data was due to inadequate funding for data preparation.

The PDS architecture is designed to provide management, analysis, and distribution services to PIs during active mission phases. PIs and Missions should be encouraged to avail themselves of the system. Costs of enhancing the basic system to meet mission requirements, i.e. high speed data links, should be borne by the mission and removed when no longer required. The elimination of the data submission task is an attractive feature of this approach. The data would be in the proprietary archive in the required format and would simply an open Investigation Node for general access.

6.4. Prioritized Accomplishment

The current fiscal climate makes it necessary to implement the PDS by prioritization. Workshop discussions resulted in the designation of three broad and overlapping categories of tasks in the implementation scheme. These are: critical, immediate, and developmental. Continuing reprioritization of tasks must accompany reviews of progress.

CRITICAL activities are:

> The saving of previous mission data and documentation in danger of being permanently lost.

Influencing ongoing planetary mission AO preparations in regard to data submission, archiving, and PI contractual agreements.

Those activities considered IMMEDIATE include:

Planning and cost analysis necessary to obtain funding for project initiation.

Communicating to the Pilot PDS those needs for items requiring long-lead development or demonstration.

Publishing the Proceedings of the Workshop and establishing continuing participation by the community.

DEVELOPMENTAL activities include all tasks leading to a full-up operational Planetary Data System. Specific among these are:

Definition of critical elements such as data standards, management structure, and virtual system.

Evaluation of new and affordable technologies.

Definition of interfaces with new and ongoing missions.

Selection of central management site, discipline centers, and data nodes, through AO process.

Development of functional elements such as data catalog, archive, network structure.

Incorporation of existing facilities and data sets into the system.

6.5. Strawman Schedule

A phased implementation of the PDS allows orderly incorporation of existing data sets and facilities into the system. The flexible schedule also permits the system to accomodate the latest developments in advancing technologies.

The PDS Implementation Schedule, Figure 6.1, shows the gradual definition and development of the PDS. The key milestones will vary as actual costs and complexity are better understood, so that the task is carried out within a fixed budget ceiling.

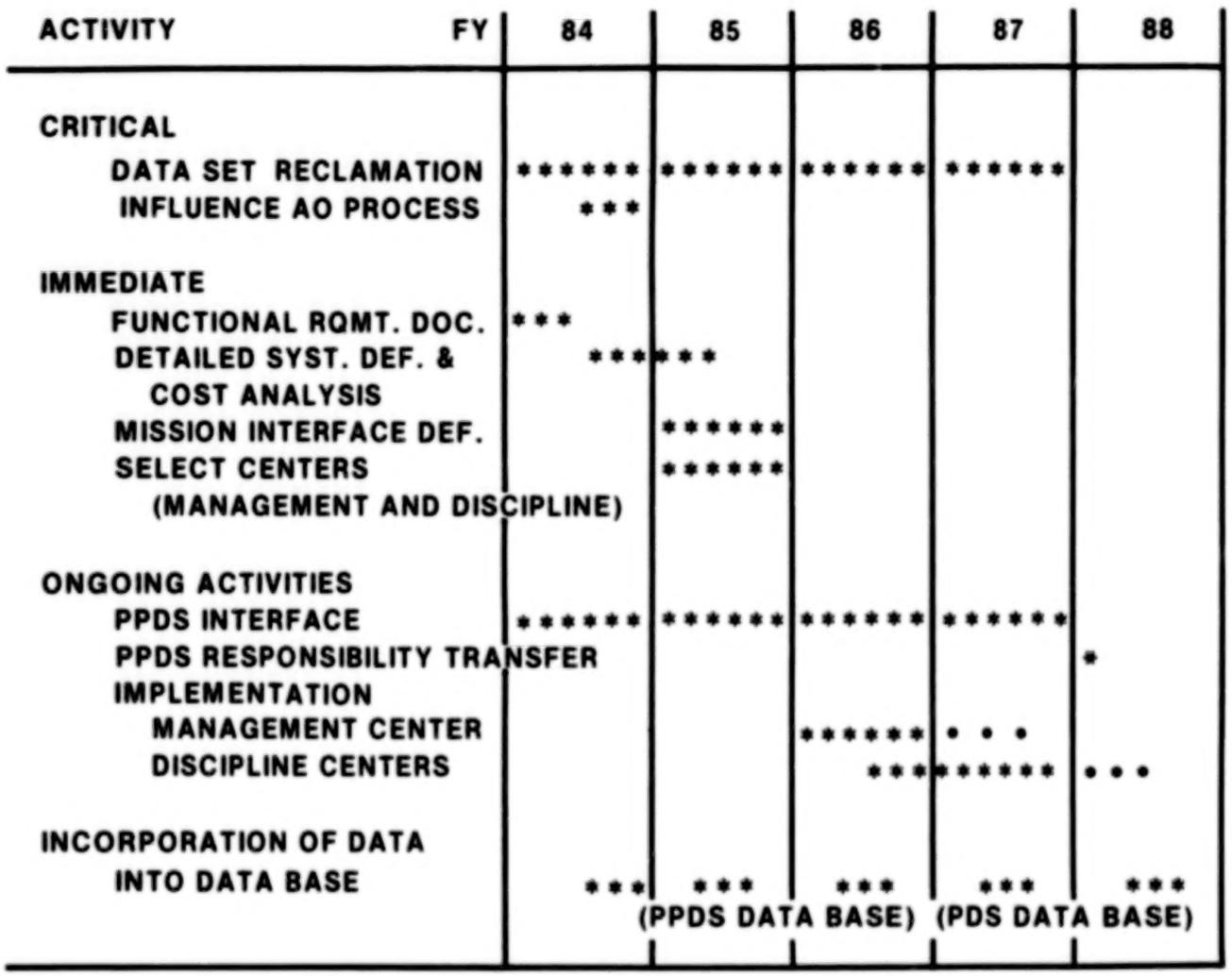

FIGURE 6.1 PDS IMPLEMENTATION SCHEDULE

7. Conclusion

Planetary Scientists envision a new way of handling data and a new way to conduct the business of performing data analysis. These ideas are born of need, frustration, and a knowledge that questions posed by the solar system may have answers in the data. However, these answers are not readily extractable using current data analysis methods.

At the Planetary Data Workshop, sponsored by NASA Code EL and held in late November of 1983, scientists considered the state of the data product and the currently used practices for finding, obtaining, and using data. They found them all to be less than satisfactory. This Executive Summary and the three-volume proceedings of that workshop constitute the functional requirements for a Planetary Data System (PDS), which scientists feel will better meet their needs for data analysis.

It is realized by all workshop participants that the implementation of a system to improve organization of and access to data, ALONE, will not solve the planetary data analysis problem. Implied were needs for better communication, standard terminologies and approaches, and a steadier supply of resources.

The Planetary Data Workshop was a unique experience and a giant step toward the definition of needs, establishment of goals, and unification of approaches to data analysis by the Planetary Science Community.

END

DATE

FILMED

NOV 30 1984